Source Coding for a Simple Multi-hop Network

Wei-Hsin Gu
Department of Electrical Engineering, MC 136-93
California Institute of Technology
1200 E. California Blvd., Pasadena, CA 91125, USA
Email: whgu@z.caltech.edu

Michelle Effros
Department of Electrical Engineering, MC 136-93
California Institute of Technology
1200 E. California Blvd., Pasadena, CA 91125, USA
Email: effros@caltech.edu

***Abstract*— We derive the rate-distortion region for source coding on a simple multihop network with side information. The result represents the first complete solution to a multihop source coding problem. The proof technique combines ideas from Wyner-Ziv coding and coding with unreliable side information.**

I. Introduction

Many of today's networking applications involve multihop networks, where a data source may be separated from its destination by one or more intermediate nodes, each of which may make its own source requests. To date, source coding theory has concentrated primarily on bounds for single-hop networks, assuming that every source has a direct connection to each destination. While single-hop network source coding solutions can be applied in multihop networks, such applications require explicit rate allocation for each source-destination pair, and the resulting solutions may be suboptimal. As a result, the study of source coding for multihop networks is an important, largely open area for investigation.

Multihop networks exhibit a variety of characteristics absent from single-hop networks: a single source description may take multiple paths to its destination; multiple source descriptions may share a single link en route to different destinations; and intermediate nodes may process incoming descriptions and send partial descriptions on to subsequent nodes in the network. The network under investigation here combines the latter two properties. To the authors' knowledge, the only prior rate-distortion theory investigations of multihop networks are Yamamoto's rate-distortion region for a single-path two-hop network without side information [1] and bounds on the rate-distortion region for a two-path multihop network [2].

Consider the network shown in Fig.1. The random sequence $(X_1, Y_1, Z_1), (X_2, Y_2, Z_2), \ldots$ is drawn i.i.d. according to joint probability mass function $p(x, y, z)$ on finite alphabet $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. The transmitter observes sources X and Y and describes them at rate R_X to the middle node. The middle node uses its received description to build a reconstruction $\hat{X}$ of source X and to create a rate-R_Y description ($R_Y \leq R_X$) for transmission to the final receiver. The final receiver combines its received description with observed side information Z to build a reconstruction $\hat{Y}$ of source Y. We measure the accuracy of reconstructions $\widehat{X}$ and $\widehat{Y}$ using distortion measures $d_X : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$ and $d_Y : \mathcal{Y} \times \hat{\mathcal{Y}} \to [0, \infty)$. We drop the subscripts from d for notational simplicity and use $d_{\max}$ to denote the maximal distortion value.

This material is based upon work partially supported by NSF Grant No. CCR-0220039 and the Caltech's Lee Center for Advanced Networking.

We define the rate-distortion region as the closure of the set of rate-distortion vectors (R_X, R_Y, D_X, D_Y) that can be achieved in coding on the given network. Our central result is a complete characterization of this rate-distortion region. The proof that follows combines ideas from Wyner-Ziv coding and coding with unreliable side information [3].

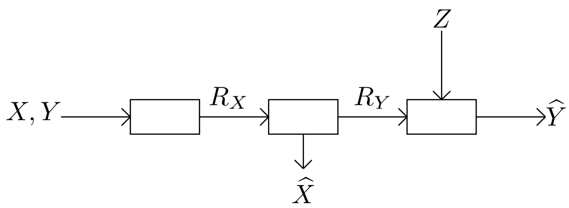

Fig. 1. The network of interest.

Theorem 1.1: *Rate-distortion vector (R_X, R_Y, D_X, D_Y) is in the rate-distortion region for lossy source coding on the network shown in Fig. 1 if and only if there exist a finite-alphabet, auxiliary random variables U and V for which*

$$\begin{aligned} R_X &\geq R_{X|U}(D_X) + I(X, Y; U) + I(X, Y; V|U, Z) \\ R_Y &\geq I(X, Y; U) + I(X, Y; V|U, Z) \end{aligned}$$

and

(i) *$Z \to (X, Y) \to (U, V)$ forms a Markov chain.*

(ii) *There exists a function $\widehat{Y}(U, V, Z)$ such that*

$$E(d(Y, \widehat{Y}(U, V, Z))) \leq D_Y.$$

(iii) *The alphabets $\mathcal{U}$ and $\mathcal{V}$ for U and V have sizes bounded by*

$$|\mathcal{U}| \leq |\mathcal{X}||\mathcal{Y}| + 3$$

and

$$|\mathcal{V}| \leq (|\mathcal{X}||\mathcal{Y}|)^2 + 3|\mathcal{X}||\mathcal{Y}| + 1.$$

The following corollary is perhaps a more intuitive characterization for the case of lossless source coding. The achievable rate region for lossless source coding is the closure of the set of rates for which we can design a sequence of codes with probability of error approaching zero.

Corollary 1.2: *Rate vector (R_X, R_Y) is in the achievable rate region for lossless source coding in the given network if and* *only if there exists finite-alphabet, auxiliary random variable U for which*

$$\begin{aligned} R_X &\geq I(X,Y;U) + H(X|U) + H(Y|U,Z) \\ R_Y &\geq I(X,Y;U) + H(Y|U,Z) \end{aligned}$$

and $Z \rightarrow (X,Y) \rightarrow U$ forms a Markov chain. Moreover, the alphabet $\mathcal{U}$ for U has size bounded by

$$|\mathcal{U}| \leq |\mathcal{X}||\mathcal{Y}| + 2.$$

Intuitively, auxiliary random variable U represents the common information that can be reconstructed at both the middle node and the final receiver without use of side information Z. Auxiliary random variable V represents the private information that can only be reconstructed at the final receiver using that receiver's knowledge of side information Z. The lossy source coding result uses the knowledge of U, V, and Z to reconstruct $\widehat{Y}$; the lossless result replaces condition (ii) with an explicit inclusion of any additional rate $H(Y|U,V,Z)$ that may be required to reconstruction Y given the knowledge of U, V, and Z.

Proof of the converse of Theorem 1.1 appears in Section II. The achievability proof appears in Section III. In addition to its direct interest, the given result can serve as a stepping stone for understanding more general multihop source coding problems.

II. The Converse

The following properties from [4] are useful in our proof of the converse of Theorem 1.1.

1) For any $\Delta \geq 0$,
$$R_{X|U}(\Delta) = \min_A \sum_u \Pr\{U = u\} R_{X|U=u}(\Delta_u), \quad (1)$$
where for each $u \in \mathcal{U}$, $R_{X|U=u}(\Delta_u)$ is the rate distortion function with respect to the probability mass function $P_{X|U=u}(x)$ on $\mathcal{X}$ and A is the set consisting of all collections $\{\Delta_u\}_{u\in\mathcal{U}}$ such that $E(\Delta_U) \leq \Delta$.
2) Given side information sources U and V, for any $\Delta \geq 0$,
$$R_{X|U,V}(\Delta) \leq R_{X|U}(\Delta). \quad (2)$$
3) Let $U_1, \ldots, U_n$ be n random variables with mutually disjoint alphabets $\mathcal{U}_1, \ldots, \mathcal{U}_n$. Let $\{\Delta_1, \ldots, \Delta_n\}$be a collection of n distortions, where $\Delta_i \geq 0$ for all i. Let $X_1, \ldots, X_n$ be drawn i.i.d. according to probability mass function $P_X(\cdot)$ on alphabet $\mathcal{X}$. Let Q be a random variable uniformly distributed on $\{1, \ldots, n\}$. Define $X := X_Q$ and $U := (U_Q, Q)$. Then
$$R_{X|U}\left(\frac{1}{n}\sum_{i=1}^n \Delta_i\right) \leq \frac{1}{n}\sum_{i=1}^n R_{X_i|U_i}(\Delta_i). \quad (3)$$

Consider a sequence $\{C_n\}_{n=1}^\infty$ of blocklength-n codes for increasing values of n. Let

$$\begin{aligned} f_n &: \mathcal{X}^n \times \mathcal{Y}^n \rightarrow \{1, \ldots, 2^{nR_X}\} \\ g_n &: \{1, \ldots, 2^{nR_X}\} \rightarrow \{1, \ldots, 2^{nR_Y}\} \end{aligned}$$

denote the rate-R_X encoder at the transmitter and rate-R_Y encoder at the middle node, respectively. Suppose that the distortions of the given codes approach D_X and D_Y as n grows without bound. Then for any $\epsilon > 0$, there exists an n sufficiently large such that the distortions achieved by code C_n are no greater than $D_X + \epsilon$ and $D_Y + \epsilon$. Let $\widehat{X}^n$ and $\widehat{Y}^n$ be the corresponding reproductions. For $i \in \{1, \ldots, n\}$, define $D_{X,i} := E[d(X_i, \widehat{X}_i)]$ and $D_{Y,i} := E[d(Y_i, \widehat{Y}_i)]$. By assumption,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n E[d(X_i, \widehat{X}_i)] &= \frac{1}{n}\sum_{i=1}^n D_{X,i} \leq D_X + \epsilon \\ \frac{1}{n}\sum_{i=1}^n E[d(Y_i, \widehat{Y}_i)] &= \frac{1}{n}\sum_{i=1}^n D_{Y,i} \leq D_Y + \epsilon. \end{aligned} \quad (4)$$

For random source sequences (X^n, Y^n), let $S = f_n(X^n, Y^n)$ and $T = g_n(S)$ represent the random variables transmitted through the first and second links respectively. Further, define $U_i := (T, X_1^{i-1}, Y_1^{i-1}, Z_1^{i-1})$ and $V_i := (Z_{i+1}^n)$. Then

$$\begin{aligned} H(T) &= I(X^n, Y^n; T) \\ &= I(X^n, Y^n; T, Z^n) - I(X^n, Y^n; Z^n|T) \\ &= \sum_{i=1}^n I(X_i, Y_i; T, Z^n | X_1^{i-1}, Y_1^{i-1}) \\ &\quad - \sum_{i=1}^n I(X^n, Y^n; Z_i | T, Z_1^{i-1}) \\ &\geq \sum_{i=1}^n I(X_i, Y_i; T, X_1^{i-1}, Y_1^{i-1}, Z_1^{i-1}, Z_i, Z_{i+1}^n) \\ &\quad - \sum_{i=1}^n I(X_i, Y_i; Z_i | T, X_1^{i-1}, Y_1^{i-1}, Z_1^{i-1}) \\ &= \sum_{i=1}^n [I(X_i, Y_i; U_i, V_i, Z_i) - I(X_i, Y_i; Z_i|U_i)] \\ &= \sum_{i=1}^n [I(X_i, Y_i; U_i) + I(X_i, Y_i; V_i, Z_i|U_i) \\ &\quad - I(X_i, Y_i; Z_i|U_i)] \\ &= \sum_{i=1}^n [I(X_i, Y_i; U_i) + I(X_i, Y_i; V_i|U_i, Z_i)]. \end{aligned}$$

The inequality follows since $(X_i, Y_i, Z_i)_{i=1}^n$ is drawn i.i.d. and (X_i, Y_i) is conditionally independent of (Z_1^{i-1}, Z_{i+1}^n) given U_i together imply

$$I(X_i, Y_i; X_1^{i-1}, Y_1^{i-1}) = 0.$$

Thus

$$nR_Y \geq H(T) \geq \sum_{i=1}^n [I(X_i, Y_i; U_i) + I(X_i, Y_i; V_i|U_i, Z_i)].$$

Now, for each $i \in \{1, 2, \ldots, n\}$, define $W_i := (T, X_1^{i-1})$. Then we have

$$\begin{aligned}
H(\widehat{X}^n|T) &\geq I(X^n; \widehat{X}^n|T) \\
&= H(X^n|T) - H(X^n|\widehat{X}^n, T) \\
&= \sum_{i=1}^{n} [H(X_i|T, X_1^{i-1}) - H(X_i|\widehat{X}^n, T, X_1^{i-1})] \\
&\geq \sum_{i=1}^{n} [H(X_i|W_i) - H(X_i|W_i, \widehat{X}_i)] \\
&= \sum_{i=1}^{n} I(X_i; \widehat{X}_i|W_i) \\
&\geq \sum_{i=1}^{n} R_{X_i|W_i}(D_{1,i}) \qquad (5) \\
&\geq \sum_{i=1}^{n} R_{X_i|U_i}(D_{1,i}), \qquad (6)
\end{aligned}$$

where (5) follows from the definition of conditional rate distortion functions and (6) follows from (2). Hence

$$\begin{aligned}
nR_X &\geq H(S) = H(S) + H(\widehat{X}^n, T|S) = H(S, \widehat{X}^n, T) \\
&\geq H(\widehat{X}^n, T) = H(\widehat{X}^n|T) + H(T) \\
&\geq \sum_{i=1}^{n} [R_{X_i|U_i}(D_{1,i}) + I(X_i, Y_i; U_i) \\
&\quad + I(X_i, Y_i; V_i|U_i, Z_i)].
\end{aligned}$$

Let Q denote a random variable uniformly distributed on $\{1, 2, \ldots, n\}$ that is independent of (X^n, Y^n, Z^n). Then from (3),

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} R_{X_i|U_i}(D_{1,i}) &\geq R_{X_Q,Q|U_Q,Q}\left(\frac{1}{n}\sum_{i=1}^{n} D_{1,i}\right) \\
&\geq R_{X_Q|U_Q,Q}(D_X + \epsilon),
\end{aligned}$$

where the last inequality follows from (4) and the fact that $R_{X_Q|U_Q,Q}(D)$ is a nonincreasing function of D. Define $U := (U_Q, Q)$ and $V := (V_Q, Q)$. Since (X_i, Y_i, Z_i), $i \in \{1, 2, \ldots, n\}$ is drawn i.i.d., the joint distribution of (X_Q, Y_Q, Z_Q) is the same as that of (X, Y, Z). Furthermore, Q is independent of (X_Q, Y_Q, Z_Q), hence

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} I(X_i, Y_i; U_i) &= I(X_Q, Y_Q; U_Q|Q) \\
&= H(X_Q, Y_Q|Q) - H(X_Q, Y_Q|U_Q, Q) \\
&= H(X_Q, Y_Q) - H(X_Q, Y_Q|U) \\
&= I(X_Q, Y_Q; U).
\end{aligned}$$

Similarly, one can show that

$$\frac{1}{n}\sum_{i=1}^{n} I(X_i, Y_i; V_i|U_i, Z_i) = I(X_Q, Y_Q; V|U, Z_Q).$$

Redefine $X = X_Q$, $Y = Y_Q$, and $Z = Z_Q$. Then X, Y, and Z have the same joint distribution $p(x, y, z)$. Therefore,

$$\begin{aligned}
R_X &\geq R_{X|U}(D_X) + I(X, Y; U) + I(X, Y; V|U, Z) \\
R_Y &\geq I(X, Y; U) + I(X, Y; V|U, Z).
\end{aligned}$$

Given the definition $(U_i, V_i) := (T, X_1^{i-1}, Y_1^{i-1}, Z_1^{i-1}, Z_{i+1}^n)$, $Z_i \to (X_i, Y_i) \to (U_i, V_i)$ forms a Markov chain, as does $Z \to (X, Y) \to (U, V)$.

It remains to check conditions (ii) and (iii) in the statement of the theorem. Since the reproduction $\widehat{Y}^n$ of Y^n is a deterministic function of (T, Z^n), we have

$$0 = H(\widehat{Y}^n|T, Z^n) = \sum_{i=1}^{n} H(\widehat{Y}_i|\widehat{Y}_1^{i-1}, T, Z^n),$$

which implies that

$$\begin{aligned}
0 &= H(\widehat{Y}_i|\widehat{Y}_1^{i-1}, T, X_1^{i-1}, Y_1^{i-1}, Z_1^{i-1}, Z_i, Z_{i+1}^n) \\
&= H(\widehat{Y}_i|\widehat{Y}_1^{i-1}, U_i, V_i, Z_i)
\end{aligned}$$

for all $i \in \{1, 2, \ldots, n\}$. For $i = 1$, $H(\widehat{Y}_1|U_1, V_1, Z_1) = 0$ implies that $\widehat{Y}_1$ is a deterministic function of (U_1, V_1, Z_1). For $i > 1$, assume that $\widehat{Y}_j$ is a function of (U_j, V_j, Z_j) for all $j < i$. Then since for all $j < i$,

$$\begin{aligned}
(U_i, V_i, Z_i) &= (T, X_1^{i-1}, Y_1^{i-1}, Z^n) \\
&= (T, X_1^{j-1}, Y_1^{j-1}, Z^n, X_j^{i-1}, Y_j^{i-1}) \\
&= (U_j, V_j, Z_j, X_j^{i-1}, Y_j^{i-1}),
\end{aligned}$$

$\widehat{Y}_j$ is also a function of (U_i, V_i, Z_i) for all $j < i$. Therefore,

$$0 = H(\widehat{Y}_i|\widehat{Y}_1^{i-1}, U_i, V_i, Z_i) = H(\widehat{Y}_i|U_i, V_i, Z_i).$$

Thus by induction on i, $\widehat{Y}_i$ is a function of (U_i, V_i, Z_i). By defining $\widehat{Y} := \widehat{Y}_Q$, $\widehat{Y}$ is a function of (U, V, Z) and

$$E(d(Y, \widehat{Y})) \leq D_Y + \epsilon.$$

Since $\epsilon > 0$ is chosen arbitrarily, (ii) holds.

Finally, we apply the support lemma in [5, p. 190] to bound the alphabet sizes $|\mathcal{U}|$, $|\mathcal{V}|$ of auxiliary random variables U and V. In bounding $|\mathcal{U}|$, we need to preserve the joint probability distribution of (X, Y), as well as $I(X, Y; U)$, $R_{X|U}(D_X)$, $I(X, Y; V|U, Z)$, and condition (ii) in the statement. Therefore, the support lemma gives

$$|\mathcal{U}| \leq |\mathcal{X}||\mathcal{Y}| - 1 + 4 = |\mathcal{X}||\mathcal{Y}| + 3.$$

In bounding $|\mathcal{V}|$, the joint probability distribution of (X, Y, U) is needed to preserve, plus two more conditions to preserve $I(X, Y; V|U, Z)$ and condition (ii). This implies that

$$\begin{aligned}
|\mathcal{V}| &\leq |\mathcal{X}||\mathcal{Y}||\mathcal{U}| + 1 \\
&\leq (|\mathcal{X}||\mathcal{Y}|)^2 + 3|\mathcal{X}||\mathcal{Y}| + 1
\end{aligned}$$

□

Remark 2.1: In the lossless case, since U is designed to preserve the joint probability distribution of (X, Y), $I(X, Y; U)$, $H(X|U)$, and $H(Y|U, Z)$, the alphabet size of U can be bounded as

$$|\mathcal{U}| \leq |\mathcal{X}||\mathcal{Y}| + 2.$$

□

III. ACHIEVABILITY

Our achievability proof relies on strongly typical sets (see, for example, [6]). Assume B is a random variable. Let $N(\beta|b^n)$ denote the number of appearances of symbol β in string b^n. We use the notation $A_\epsilon^{*(n)}(B)$ to denote the strongly typical set for random variable B on alphabet $\mathcal{B}$, where $A_\epsilon^{*(n)}(B)$ is the set of sequences $b^n \in \mathcal{B}^n$ satisfying

(a)
$$\left|\frac{N(\beta|b^n)}{n} - p(\beta)\right| < \frac{\epsilon}{|\mathcal{B}|}$$
for every $\beta \in \mathcal{B}$ with $p(\beta) > 0$.

(b) $N(\beta|b^n) = 0$ for all $\beta \in \mathcal{B}$ with $p(\beta) = 0$.

If B is another random variable and $c^n \in A_\epsilon^{*(n)}(C)$, define
$$A_\epsilon^{*(n)}(B|c^n) := \{b^n \in B^n | (b^n, c^n) \in A_\epsilon^{*(n)}(B, C)\}.$$
The following 2 lemmas show some properties of typical sets that are useful in proving the achievability result that follows.

Lemma 3.1: For any random variables $W, \widehat{W}$ with bounded distortion measure $d(\cdot,\cdot)$ and every strongly jointly typical pair $(w^n, \widehat{w}^n) \in A_\epsilon^{*(n)}(W)$,
$$\begin{aligned}
&\left|\frac{1}{n}d(w^n, \widehat{w}^n) - Ed(W, \widehat{W})\right| \\
&= \left|\sum_{\alpha\in\mathcal{W},\beta\in\widehat{\mathcal{W}}} \left(\frac{N(\alpha,\beta|w^n,\widehat{w}^n)}{n} - p(\alpha,\beta)\right) d(\alpha,\beta)\right| \\
&\le d_{\max} \sum_{\alpha\in\mathcal{W},\beta\in\widehat{\mathcal{W}}} \frac{\epsilon}{|\mathcal{W}||\widehat{W}|} \\
&= \epsilon \cdot d_{\max}.
\end{aligned}$$
□

In the next lemma, one can easily see that 2) and 3) are direct consequences of 1) that estimates the sizes of conditional strongly typical sets.

Lemma 3.2: 1) *[6, Problem 13.10] For any $\epsilon > 0$, if $x^n \in A_\epsilon^{*(n)}(X)$, then*
$$2^{n(H(Y|X)-\epsilon')} \le |A_\epsilon^{*(n)}(Y|x^n)| \le 2^{n(H(Y|X)+\epsilon')},$$
where ϵ' can be made arbitrarily small by making n sufficiently large and ϵ sufficiently small.

2) *Given a probability distribution $p(x, y, w)$, fix any pair $(x^n, w^n) \in A_\epsilon^{*(n)}(X, W)$, and choose a sequence Y^n uniformly at random from the set $A_\epsilon^{*(n)}(Y|w^n)$. Then*
$$\begin{aligned}2^{-n(I(X;Y|W)+\epsilon_1)} &\le \Pr(Y^n \in A_\epsilon^{*(n)}(Y|(x^n, w^n)) \\ &\le 2^{-n(I(X;Y|W)-\epsilon_1)},\end{aligned}$$
where ϵ_1 can be made arbitrarily small by making n sufficiently large and ϵ sufficiently small.

3) *Given a probability distribution $p(x, y, w)$ and a typical sequence w^n, independently choose $2^{\lfloor nR\rfloor}$ sequences $Y_1^n, Y_2^n, \ldots, Y_{2^{\lfloor nR\rfloor}}^n$ from the set $A_\epsilon^{*(n)}(Y|w^n)$. When $R > I(X;Y|W)$,*
$$\Pr((X^n, Y_i^n, w^n) \in A_\epsilon^{*(n)}(X, Y, W) \text{ for some } i \in \{1, 2, \ldots, 2^{\lfloor nR\rfloor}\}) \to 1$$
as $n \to \infty$.

4) *[6, **Lemma 14.8.1**] Let $X \to Y \to Z$ form a Markov chain. If for a given $(y^n, z^n) \in A_\epsilon^{*(n)}(Y, Z)$, X^n is chosen uniformly at random from the set of x^n that are jointly typical with y^n, then*
$$Pr\{X^n \in A_\epsilon^{*(n)}(X|y^n, z^n)\} > 1 - \epsilon$$
for sufficiently large n.

□

Let (U, V) be a pair of random variables satisfying conditions (i) and (ii) In Theorem 1.1. To prove achievability, it suffices to show that for any $\delta > 0$, the vector $(R_X(\delta), R_Y(\delta))$ defined by
$$\begin{aligned}R_X(\delta) &= R_{X|U}(D_X) + I(X,Y;U) + I(X,Y;V|U,Z) + 3\delta \\ R_Y(\delta) &= I(X,Y;U) + I(X,Y;V|U,Z) + 2\delta\end{aligned}$$
is achievable.

Fix $n \in \mathbb{N}$. Let $\widehat{X}$ be a random variable defined by a conditional probability distribution $p(\widehat{x}|x, u)$ such that $I(X;\widehat{X}|U) = R_{X|U}(D_X) + \delta/2$ and $E(d(X,\widehat{X})) \le D_X$. Define
$$\begin{aligned}S &:= 2^{\lfloor n(I(X,Y;V|U)+\delta)\rfloor} & M &:= 2^{\lfloor n(I(X,Y;U)+\delta)\rfloor} \\ T &:= 2^{\lfloor n(I(X;\widehat{X}|U)+\delta)\rfloor} & N &:= 2^{\lfloor n(I(X,Y;V|U,Z)+\delta)\rfloor}.\end{aligned}$$
Generate the codebook as follows:

1) Randomly choose M sequences $U^n(1),\ldots,U^n(M)$ i.i.d. according to probability mass function $\prod_{i=1}^n p(u_i)$.
2) For each $m \in \{1,\ldots,M\}$, randomly choose S sequences $V^n(m,1),\ldots,V^n(m,S)$ uniformly at random from the set of $v^n \in \mathcal{V}^n$ that are strongly jointly typical with $U^n(m)$.
3) For each $m \in \{1,\ldots,M\}$, randomly choose T sequences $\widehat{X}^n(m,1),\ldots,\widehat{X}^n(m,T)$ uniformly at random from the set of $x^n \in \mathcal{X}^n$ that are strongly jointly typical with $U^n(m)$.

Let (x^n, y^n) and z^n be the source pair to be transmitted and the observed side information, respectively. For each $j \in \{1,\ldots,S\}$, draw $\tau(j)$ uniformly at random from $\{1, 2,\ldots,N\}$. We use the following functions in defining the encoders
$$\begin{aligned}
\psi(x^n, y^n) &:= \min\Big[\{M\} \cup \{m \in \{1,\ldots,M\} : \\ &\qquad (x^n, y^n, U^n(m)) \in A_\epsilon^{*(n)}(X, Y, U)\}\Big]. \\
\mu(x^n, y^n) &:= \min[\{S\} \cup \{j \in \{1,\ldots,S\} : \\ &\qquad (x^n, U^n(\psi(x^n, y^n)), V^n(\psi(x^n, y^n), j)) \\ &\qquad \in A_\epsilon^{*(n)}(X, U, V)\}\Big] \\
\phi(x^n, y^n) &:= \min[\{T\} \cup \{t \in \{1, 2,\ldots,T\} : \\ &\qquad (x^n, U^n(\psi(x^n, y^n)), \widehat{X}^n(\psi(x^n, y^n), t)) \\ &\qquad \in A_\epsilon^{*(n)}(X, U, \widehat{X})\}\Big] \\
\pi(x^n, y^n) &:= \tau(\mu(x^n, y^n)).
\end{aligned}$$
The purpose of using the function τ is that, while S is the number of sequences in V that can cover all strongly typical

pairs in (X, Y) knowing U, we further randomly bin the index j indicating the desired sequence in V into one of the N slots. Since the decoder has access to the side information Z, one hopes that it is possible to recover j from $\tau(j)$ plus the side information Z with sufficiently high probability.

Finally, we define the encoders as

$$\begin{aligned} f(x^n, y^n) &:= (\phi, \psi, \pi)(x^n, y^n) \\ g(f(x^n, y^n)) &:= (\psi, \pi)(x^n, y^n). \end{aligned}$$

The decoder at the middle node maps index (t, m, p), the received string from the encoding function f, to reconstruction

$$\widehat{X}_j^n(m, t).$$

At the end node, find s from $\{1, \ldots, S\}$ such that

$$\begin{aligned} &(U^n(m), V^n(m, s), Z^n) \in A_\epsilon^{*(n)}(U, V, Z) \\ &\tau(s) = p. \end{aligned} \tag{7}$$

Define the function ι mapping from $\{1, \ldots, M\} \times \{1, \ldots, N\}$ to $\{1, \ldots, S\}$ as follows

$$\iota(m, p) := \begin{cases} s, & \text{if } s \text{ is unique such that (7) holds} \\ 1, & \text{otherwise.} \end{cases}$$

Reproduce Y^n as

$$\widehat{Y}^n(U^n(m), V^n(m, \iota(m, p)), Z^n).$$

Analysis of performance:
For simplicity, we denote by ψ, μ, ϕ, ι, and π the evaluated values of the corresponding functions on (x^n, y^n). Define the following atypicality events:

1) $E_0 : (x^n, y^n, z^n) \notin A_\epsilon^{*(n)}(X, Y, Z)$.
2) E_U : For all $i \in \{1, 2, \ldots, M\}$ and for all $j \in \{1, 2, \ldots, T\}$,
$$(x^n, y^n, z^n, U^n(i), V^n(i, j)) \notin A_\epsilon^{*(n)}(X, Y, Z, U, V).$$
3) E_X : For all $i \in \{1, 2, \ldots, T\}$,
$$(x^n, U^n(\psi), \widehat{X}^n(\psi, i)) \notin A_\epsilon^{*(n)}(X, U, \widehat{X}).$$
4) $E_V : (U^n(\psi), z^n) \in A_\epsilon^{*(n)}(U, Z)$ and there exists $V^n(\psi, j) \neq V^n(\psi, \mu)$ such that
$$(U^n(\psi), V^n(\psi, j), z^n) \in A_\epsilon^{*(n)}(U, V, Z)$$
and $\tau(j) = \pi$.
5) $E = E_0 \cup E_U \cup E_X \cup E_V$.

From the basic property of typical sets, $\Pr(E_0)$ can be made arbitrarily small as n grows without bound. By *Lemma 3.2 3)* and 4), $\Pr(E_U)$ and $\Pr(E_X)$ can also be made arbitrarily small as n grows without bound. Finally, when $(U^n(\psi), z^n) \in A_\epsilon^{*(n)}(U, Z)$, since the probability that a randomly chosen sequence $v^n \in A_\epsilon^{*(n)}(V|U^n(\psi))$ is jointly strongly typical with $(U^n(\psi), z^n)$ is approximately $2^{nI(V;Z|U)}$, the probability of the event E_V can be bounded (for n large enough) by

$$\Pr(E_V) \leq |B(p)| 2^{n(I(V;Z|U) - \frac{\delta}{2})} = \frac{S}{N} \times 2^{n(I(V;Z|U) - \frac{\delta}{2})},$$

where $B(p)$ is the set of indices $j \in \{1, \ldots, S\}$ such that $\tau(j) = p$. Since

$$\begin{aligned} &I(X, Y; V|U) - I(X, Y; V|U, Z) \\ &\quad = (H(V|U) - H(V|U, X, Y)) \\ &\quad\quad - (H(V|U, Z) - H(V|X, Y, U, Z)) \\ &\quad = H(V|U) - H(V|U, Z) \\ &\quad\quad + (H(V|X, Y, U, Z) - H(V|U, X, Y)) \\ &\quad = H(V|U) - H(V|U, Z) = I(V; Z|U), \end{aligned}$$

$\Pr(E_V)$ can also be made arbitrarily small as n grows without bound.
Hence by the union bound, $\lim_{n\to\infty} \Pr(E) = 0$. From Lemma 3.1, the average distortion between any pair of jointly typical sequences is close to the expected distortion. Therefore,

$$\begin{aligned} &|Ed(X^n, \widehat{X}^n(\psi, \phi)) - D_X| \\ &= \Big| \Pr(E^c) \left[E\left(d(X^n, \widehat{X}^n(\psi, \phi) | E^c \right) - D_X \right] \\ &\quad + \Pr(E) \left[E\left(d(X^n, \widehat{X}^n(\psi, \phi) | E \right) - D_X \right] \Big| \\ &\leq \Pr(E^c) \cdot \epsilon \cdot d_{\max} + \Pr(E) \cdot d_{\max} \\ &\leq \epsilon \cdot d_{\max} + \Pr(E) \cdot d_{\max}. \end{aligned}$$

Similarly,

$$\begin{aligned} &|Ed(Y^n, \widehat{Y}^n(U^n(\psi), V^n(\psi, \iota(\psi, \tau)), Z^n) - D_Y| \\ &< \epsilon \cdot d_{\max} + \Pr(E) \cdot d_{\max}. \end{aligned}$$

Since $\epsilon > 0$ is chosen arbitrarily and $\Pr(E)$ can be made arbitrarily small for n sufficiently large, this coding scheme satisfies the distortion requirement (D_X, D_Y). This completes the proof.

References

[1] H. Yamamoto, "Source Coding Theory for Cascade and Branching Communication Systems," *IEEE Trans. on Information Theory,* IT-27, May 1981, pp. 299–308.

[2] W. H. Gu and M. Effros, "Source Coding for a Multi-Hop Network," IEEE Data Compression Conference, Snowbird, Utah, March 2005.

[3] C. Heegard and T. Berger, "Rate Distortion when Side Information May Be Absent," *IEEE Trans. on Informaiton Theory,* IT-31, pp. 727–734, Nov. 1985.

[4] R. M. Gray, "A New Class of Lower Bounds to Information Rates of Stationary Sources via Conditional Rate-Distortion Functions," *IEEE Trans. on Information Theory,* IT-19,July 1973, pp. 480–489.

[5] I. Csiszár and J. Köner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. London, U.K.:Academic, 1981.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience Publication, 1991.